



MEDFAMYC

REVISTA DE MEDICINA FAMILIAR Y COMUNIDAD
UNIVERSIDAD NACIONAL DE ASUNCIÓN

ISSN: 3105-9899 | Vol. 3 Núm. 1 | Enero-junio, 2026

PERSPECTIVA

La «empatía sin sujeto»: Límites bioéticos de la inteligencia artificial afectiva en la relación médico-paciente de atención primaria

Andrea Paola Britos Gómez¹, Alcides Chaux¹

¹ChauxLab Institute, Asunción, Paraguay

DOI: [10.66476/mf.v3n1.7185](https://doi.org/10.66476/mf.v3n1.7185)



RESUMEN

Introducción: La inteligencia artificial (IA) puede funcionar como una «tercera parte» capaz de aliviar la carga burocrática de la atención primaria de salud (APS) y, con ello, devolver tiempo a la relación médico-paciente. Eso deja abierta una pregunta que este artículo retoma: si la IA se aproxima cada vez más a producir expresiones que los pacientes perciben como empáticas, ¿qué estatuto ético tiene esa empatía cuando no existe, detrás de ella, un sujeto que sienta? **Objetivo:** Examinar críticamente los límites bioéticos de la empatía simulada por sistemas de IA conversacional en la consulta de medicina familiar, distinguiendo su función legítima como andamiaje comunicacional de su riesgo como sustituto afectivo. **Métodos:** Se examinan de forma crítica los hallazgos recientes de la literatura en filosofía de la mente, psicología social, ética clínica y salud digital (2020-2026), en particular los estudios que comparan empatía humana y generada por IA. **Resultados y discusión:** Evaluadores y pacientes tienden a calificar las respuestas generadas por IA como más compasivas y empáticas que las de médicos humanos, incluso cuando se revela su origen artificial. Sin embargo, cuando se les permite elegir la fuente de una expresión empática, las personas prefieren recibirla de un ser humano: la empatía artificial parece funcionalmente superior pero relacionamente insuficiente. Este desacople es el núcleo del problema bioético: la IA simula con fidelidad los componentes cognitivo y comunicacional de la empatía, pero —según el análisis desarrollado en este artículo— carecería del componente motivacional, el interés genuino por el bienestar del otro, que la sostiene como fenómeno moral. El riesgo clínico es la «empatía decepiva»: respuestas que generan una sensación de conexión emocional sin comprensión ni compromiso real, capaces de inducir una alianza terapéutica digital mal fundamentada, sobre todo en pacientes vulnerables. **Conclusión:** La empatía artificial tiene un lugar legítimo como andamiaje comunicacional, pero no puede ocupar el lugar de la empatía clínica sin comprometer la confianza fiduciaria. Su incorporación exige

Enviado: 20 de mayo, 2026 | **Aceptado:** 20 de junio, 2026 | **Publicado:** 26 de junio, 2026

Correspondencia: Alcides Chaux. ChauxLab Institute, Asunción, Paraguay. **Correo electrónico:** alcides.chaux@chauxlab.com.

Financiación: Los autores declaran que no recibieron fondos externos para la realización de este estudio.

Conflictos de interés: Los autores declaran no tener conflictos de interés financieros, personales, profesionales, institucionales ni de otra índole relacionados con este trabajo.

transparencia sobre el origen de la respuesta, consentimiento informado específico y salvaguardas reforzadas en contextos de recursos limitados.

Palabras clave: inteligencia artificial, empatía, bioética, relación médico-paciente, medicina familiar, consentimiento informado.

INTRODUCCIÓN

La inteligencia artificial (IA), lejos de ser necesariamente una fuerza deshumanizante, posee el potencial de rehumanizar la atención primaria de salud (APS) al absorber la carga administrativa que hoy consume buena parte del tiempo médico (1). Ese planteamiento se centra, en todo caso, en la capacidad de la IA para automatizar tareas. Sin embargo, existe una función distinta y cada vez más presente, que es el punto de partida de este artículo: la IA que interviene en el tono afectivo de la comunicación clínica, que redacta mensajes al paciente o que simula, con creciente sofisticación, las señales lingüísticas de la comprensión y el cuidado.

Esta función es cada vez más difícil de ignorar. Los modelos de lenguaje de gran escala generan respuestas que evaluadores humanos, incluyendo profesionales de salud licenciados, califican sistemáticamente como más empáticas que las de los propios médicos (2,3), en un contexto donde la propia relación entre eficiencia, empatía y confianza ya venía tensionada antes de la IA generativa (4). Este hallazgo, replicado en distintos diseños y poblaciones, obliga a una pregunta que la bioética médica no puede eludir: ¿qué se mide exactamente cuando se mide la «empatía» de un sistema que, por definición, no experimenta nada? Este artículo examina esa pregunta desde tres ángulos: qué imita la IA cuando produce lenguaje empático y qué no puede imitar en principio; cómo perciben pacientes y evaluadores esa imitación; y qué salvaguardas exige su incorporación a la medicina familiar, con atención al eje de equidad que atraviesa este análisis.

LO QUE LA IA IMITA Y LO QUE NO PUEDE IMITAR

La empatía integra al menos tres componentes: uno cognitivo (reconocer el estado emocional del otro), uno experiencial (compartir en algún grado ese estado) y uno motivacional (el interés genuino en aliviar el sufrimiento ajeno) (5). Los sistemas conversacionales aproximan razonablemente el primero y producen, a partir de él, un lenguaje que imita con fluidez el segundo. El componente que resulta más difícil de replicar, según un análisis bioético influyente sobre los obstáculos de principio para la IA empática, es el tercero: la motivación genuina de reducir el sufrimiento de otro ser, ausente en un sistema sin sujeto experiencial (6).

Ese mismo análisis sostiene que la reciprocidad afectiva —el hecho de que ambas partes sean sujetos capaces de sentir y de ser afectados— es constitutiva, no accesorio, del efecto terapéutico de la empatía clínica (6). Otros autores matizan esta posición: precisamente porque la IA no sufre fatiga ni sesgo intergrupar, sus expresiones empáticas pueden ser más consistentes que la empatía humana intermitente (7). El desacuerdo persiste entre quienes sostienen que la IA nunca transmitirá la esencia de la conexión humana (8) y quienes defienden que su valor funcional puede justificarse con independencia de su origen ontológico.

LA PARADOJA DE LA PREFERENCIA

En un estudio ya clásico, un panel de profesionales de salud evaluó respuestas de un chatbot y de médicos a preguntas reales de pacientes: las respuestas de IA fueron calificadas como empáticas casi diez veces más que las médicas (2). Hallazgos posteriores muestran el mismo patrón frente a personal experto en crisis, incluso cuando se revela el origen artificial de la respuesta (9). Sin embargo, cuando la misma respuesta se atribuye a un humano se percibe como más empática que cuando se atribuye a la IA (10): el origen atribuido, y no solo el contenido, parece incidir en el efecto relacional.

Un hallazgo particularmente revelador introduce una variable ausente en estudios previos: la posibilidad de elegir. Cuando se ofrece a los participantes optar entre recibir apoyo emocional de un humano o de una IA, la mayoría elige al humano, pese a calificar las respuestas de IA como de mayor calidad (11). Sus autores llaman a esto la «paradoja de elección de la empatía de IA»: las personas se benefician de ella cuando la reciben, pero no la eligen

cuando tienen alternativa (11). Este hallazgo sugiere que la preferencia por lo humano no responde solo a la calidad de la respuesta, sino también a un juicio implícito sobre la legitimidad de su fuente.

LA EMPATÍA DECEPIVA

Uno de los riesgos bioéticos más relevantes no es la empatía artificial declarada como tal, sino la que se presenta —por diseño o por expectativa del usuario— como equivalente a la humana. Un marco reciente construido con psicólogos clínicos identifica la «empatía decepiva» como una de las principales violaciones éticas en consejería asistida por IA: respuestas que simulan conexión emocional sin comprensión contextual ni compromiso real (12), un patrón que revisiones sistemáticas más amplias sobre la ética de estos modelos en medicina describen también como recurrente (13). Este riesgo se documenta asimismo como «malinterpretación terapéutica»: el diseño mismo del sistema —disponibilidad permanente, tono cálido y no enjuiciador— puede llevar al usuario a sobreestimar la capacidad del chatbot de sostener una relación terapéutica genuina, en lo que se ha denominado una «alianza terapéutica digital» cuya solidez conceptual y cuyas implicancias éticas de fondo —personalización, consentimiento, dependencia a largo plazo— siguen sin resolverse (14,15). En medicina familiar el riesgo no se limita a chatbots de salud mental: cualquier sistema que medie la comunicación entre consultas puede adoptar, sin proponérselo, un registro afectivo que el paciente interprete como proveniente del médico, lo cual puede vulnerar el principio de no maleficencia (16).

CONFIANZA, CONSENTIMIENTO Y EQUIDAD

El consentimiento informado debe evolucionar para explicitar el uso de IA en los procesos diagnósticos y terapéuticos (1). La empatía simulada precisa esa exigencia: no basta informar que un sistema participó en un mensaje; conviene distinguir su uso instrumental —documentación, apoyo diagnóstico— de su uso relacional, diseñado para transmitir comprensión. Un análisis bioético sobre la IA conversacional en psicoterapia sitúa el riesgo en la ambigüedad no resuelta entre herramienta y agente: sos-

tiene que el sistema no es un agente pleno, pero tampoco una herramienta neutra, porque está optimizado para producir la impresión de reciprocidad (17). La guía de la OMS sobre ética y gobernanza de la IA en salud responde a esto desde el principio de transparencia: el paciente tiene derecho a saber cuándo interactúa con IA y a comprender sus límites (18), un derecho que cobra relieve porque las actitudes de los pacientes hacia la IA diagnóstica y comunicacional varían de forma marcada según su origen sociocultural (19).

Este principio pesa más en países de ingresos bajos y medianos, donde la escasez de personal de salud mental capacitado vuelve atractiva la promesa de una IA empática siempre disponible (20). El riesgo para la equidad algorítmica en general, ya señalado (1), es que el acceso a un profesional humano se convierta en privilegio de quienes pueden costearlo, mientras las poblaciones de menores recursos quedan circunscritas a su simulación —un riesgo de segregación afectiva que la justicia distributiva no debería ignorar (21).

CONCLUSIÓN

La IA puede producir expresiones que pacientes y evaluadores perciben como empáticas, y en ciertos contextos las evalúan por encima de las de un médico humano. Esta capacidad no equivale, sostenemos aquí, a poseer empatía en el sentido moralmente relevante: la ausencia de un sujeto motivado por el bienestar del otro es, en nuestra lectura, una limitación de principio y no solo un problema técnico que resolverían futuras versiones del modelo —aunque, como se discutió más arriba, esta sigue siendo una posición en debate. La empatía artificial tiene un lugar legítimo como andamiaje que amplifica la comunicación del médico de familia —no que la sustituye—, de la misma manera en que un escriba digital asiste la documentación sin reemplazar el juicio clínico. Ese lugar exige transparencia sobre el origen de cada interacción, consentimiento informado específico para sus usos relacionales y vigilancia reforzada en los contextos de recursos limitados donde la asimetría entre lo que el sistema simula y lo que el paciente puede verificar es mayor. Solo así la empatía simulada podrá sostener, en lugar de sustituir, el encuentro humano que sigue siendo el centro de la atención primaria de salud.

REFERENCIAS

1. Chauv A. La «tercera parte» en la consulta: Inteligencia artificial y la rehumanización de la relación médico-paciente en atención primaria. *Rev Med Fam Comunidad*. 2025;2(2):63-66. <https://doi.org/10.66476/mf.v2n2.6275>

2. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Intern Med.* 2023;183(6):589-596. <https://doi.org/10.1001/jamainternmed.2023.1838>
3. Ruben MA, Blanch-Hartigan D, Hall JA. What is artificial intelligence (AI) "empathy"? A study comparing ChatGPT and physician responses on an online forum. *J Gen Intern Med.* 2026;41:1304-1311. <https://doi.org/10.1007/s11606-025-10068-w>
4. Kerasidou A. Artificial intelligence and the ongoing need for empathy, compassion and trust in healthcare. *Bull World Health Organ.* 2020;98(4):245-250. <https://doi.org/10.2471/BLT.19.237198>
5. Decety J, Fotopoulou A. Why empathy has a beneficial impact on others in medicine: unifying theories. *Front Behav Neurosci.* 2014;8:457. <https://doi.org/10.3389/fnbeh.2014.00457>
6. Montemayor C, Halpern J, Fairweather A. In principle obstacles for empathic AI: why we can't replace human empathy in healthcare. *AI Soc.* 2022;37(4):1353-1359. <https://doi.org/10.1007/s00146-021-01230-z>
7. Inzlicht M, Cameron CD, D'Cruz J, Bloom P. In praise of empathic AI. *Trends Cogn Sci.* 2024;28(2):89-91. <https://doi.org/10.1016/j.tics.2023.12.003>
8. Perry A. AI will never convey the essence of human empathy. *Nat Hum Behav.* 2023;7(11):1808-1809. <https://doi.org/10.1038/s41562-023-01675-w>
9. Ovsyannikova D, Oldemburgo de Mello V, Inzlicht M. Third-party evaluators perceive AI as more compassionate than expert humans. *Commun Psychol.* 2025;3:4. <https://doi.org/10.1038/s44271-024-00182-6>
10. Rubin M, Li JZ, Zimmerman F, Ong DC, Goldenberg A, Perry A. Comparing the value of perceived human versus AI-generated empathy. *Nat Hum Behav.* 2025;9(11):2345-2359. <https://doi.org/10.1038/s41562-025-02247-w>
11. Wenger JD, Cameron CD, Inzlicht M. People choose to receive human empathy despite rating AI empathy higher. *Commun Psychol.* 2026;4(1):19. <https://doi.org/10.1038/s44271-025-00387-3>
12. Ifthikhar Z, Xiao A, Ransom S, Huang J, Suresh H. How LLM counselors violate ethical standards in mental health practice: a practitioner-informed framework. *Proc AAAI/ACM Conf AI Ethics Soc.* 2025;8(2):1311-1323. <https://doi.org/10.1609/aies.v8i2.36632>
13. Haltaufderheide J, Ranisch R. The ethics of ChatGPT in medicine and healthcare: a systematic review on Large Language Models (LLMs). *NPJ Digit Med.* 2024;7(1):183. <https://doi.org/10.1038/s41746-024-01157-x>
14. Khawaja Z, Bélisle-Pipon JC. Your robot therapist is not your therapist: understanding the role of AI-powered mental health chatbots. *Front Digit Health.* 2023;5:1278186. <https://doi.org/10.3389/fdgth.2023.1278186>
15. Malouin-Lachance A, Capolupo J, Laplante C, Hudon A. Does the digital therapeutic alliance exist? Integrative review. *JMIR Ment Health.* 2025;12:e69294. <https://doi.org/10.2196/69294>
16. Coghlan S, Leins K, Sheldrick S, Cheong M, Gooding P, D'Alfonso S. To chat or bot to chat: Ethical issues with using chatbots in mental health. *Digit Health.* 2023;9:20552076231183542. <https://doi.org/10.1177/20552076231183542>
17. Sedlakova J, Trachsel M. Conversational artificial intelligence in psychotherapy: a new therapeutic tool or agent? *Am J Bioeth.* 2023;23(5):4-13. <https://doi.org/10.1080/15265161.2022.2048739>
18. World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance. Geneva: World Health Organization; 2021. Disponible en: <https://iris.who.int/server/api/core/bitstreams/f780d926-4ae3-42ce-a6d6-e898a5562621/content>
19. Robertson C, Woods A, Bergstrand K, Findley J, Balser C, Slepian MJ. Diverse patients' attitudes towards Artificial Intelligence (AI) in diagnosis. *PLOS Digit Health.* 2023;2(5):e0000237. <https://doi.org/10.1371/journal.pdig.0000237>
20. Goodman KW, Litewka SG, Malpani R, Pujari S, Reis AA. Global health and big data: The WHO's artificial intelligence guidance. *S Afr J Sci.* 2023;119(5/6). <https://doi.org/10.17159/sajs.2023/14725>
21. Knox B, Christoffersen P, Leggitt K, Woodruff Z, Haber MH. Justice, vulnerable populations, and the use of conversational AI in psychotherapy. *Am J Bioeth.* 2023;23(5):48-50. <https://doi.org/10.1080/15265161.2023.2191040>